Postal Regulatory Commission
Submitted 7/22/2015 4:53:02 PM
Filing ID: 92881
Accepted 7/23/2015

BEFORE THE
POSTAL REGULATORY COMMISSION
WASHINGTON, DC 20268-0001

PERIODIC REPORTING
(PROPOSAL 13)

Docket No. RM2015-7

SECOND SUPPLEMENTAL REPORT OF
KEVIN NEELS
ON BEHALF OF
UNITED PARCEL SERVICE

Table of Contents

I. Purpose of This Report

On April 23, 2015 the Postal Regulatory Commission ("the Commission") ordered the Postal Service to provide UPS with access to a confidential crosswalk file identifying the true zip codes associated with the entries in the National Form 3999 Data Set ("the Form 3999 Data Set"). This information permitted consultants working with UPS to explore the feasibility of using the Form 3999 data to develop an improved method for calculating the variable costs associated with the Postal Service's city delivery operations. On June 8, 2015 I filed a report (the "June Report") on behalf of UPS describing a flexible form econometric model developed using the Form 3999 Data Set. This model related total gross street time in a zip code to a vector of variables describing the volume and mix of mail being delivered, and the characteristics of the delivery environment. On July 8 the Postal Service, Amazon and other parties filed comments on my June Report. Among the materials filed were two reports prepared by outside academic experts – one by Professor Christian T. Lundblad of the University of North Carolina, filed on behalf of Amazon (the "Lundblad Report"), and another by Professor Michael D. Bradley of George Washington University, filed on behalf of the Postal Service (the "Bradley Report"). The Commission has provided me and UPS with an opportunity to respond to the comments filed on July 8th. This report contains my response.

II. Summary of My Findings

The Commission should not reject the National Form 3999 Model simply because it relies upon imputed measures of volume for certain mail streams. Reliance on imputations was an element of the plan for the development of this model that was originally presented to the Commission. After having permitted UPS to proceed with the implementation of this plan, the Commission should consider its results on their merit and base its decision on their reasonableness and reliability.

I have addressed many of the criticisms of the National Form 3999 Model offered in reports produced on behalf of the Postal Service and Amazon. In particular, I have made a number of changes designed to improve the out of sample forecasting properties of the imputation models used in this analysis, to better capture changes in mail volumes over the period spanned by the Form 3999 Data Set, to reduce collinearity with other independent variables in the model, and to

reduce the model's reliance on the DOIS parcel volume data whose reliability has been so vigorously challenged by the Postal Service and other parties to this proceeding.

As a result of these improvements to the imputation models presented in my June Report they have better forecasting properties and pass validity checks based upon independently collected nationwide data.

These improved imputation results, in combination with other refinements, have resulted in an updated and more robust National Form 3999 Model whose results are consistent with available benchmarks put forward by the Postal Service.

These refinements have only strengthened the original advantages offered by the National Form 3999 Model. It continues to offer the possibility of larger sample sizes, greater degrees of statistical precision, and more frequent updates. Refinements to the model specification have simplified the task of collecting more accurate parcel volume data that could in a relatively short period of time dispense with the need to rely on imputed volume measures.

The Commission should adopt the National Form 3999 Model for use in the FY2015 Annual Compliance Determination recomputing marginal costs and variabilities as needed to reflect volumes and conditions in that year. The Commission should also direct the Postal Service to collect more accurate parcel volume data as part of its regular route evaluation process.

The various criticisms that have been offered of the Modified Proposal Thirteen discussed in my June Report are neither conceptually nor statistically persuasive. If the Commission does not adopt the National Form 3999 Model, it should adopt Modified Proposal Thirteen for use in the FY2015 Annual Compliance Determination.

III. Comments on the Lundblad Report

Professor Lundblad makes a number of comments relating to the National Form 3999 Model and the results of earlier analyses I have reported. His primary focus seems to be upon the problem of multicollinearity. I first address his comments on this subject, and then respond to his other comments and observations.

A. MULTICOLLINEARITY

Professor Lundblad focuses much of his report on the problem of multicollinearity. In its opening pages he states:

> "multicollinearity – a high degree of correlation between two or more explanatory variables – can limit the ability to estimate the effect of individual explanatory variables on the dependent variable. This is a critical concern, and much, but not all, of my review of the UPS-proposed models focuses on this issue."[1]

True to his word, he returns to this subject repeatedly over the course of his report.

> "The primary flaw in the National Form 3999 Model is its use of imputed parcel and collection mail volumes. Substantial multicollinearity results, rendering an assessment of marginal costs impossible."[2]

> "When multicollinearity is present, even modest changes in the data can produce wide swings in the resulting coefficient estimates, wreaking havoc on the analysis and limiting the conclusions one can draw from it."[3]

> "Multicollinearity and poor data quality prevent Dr. Neels' Modified Proposal 13 Model from establishing any reliable relationship between parcel volume and regular delivery time."[4]

> "The presence of this substantial multicollinearity precludes any meaningful inference about relative coefficient magnitudes (let alone any reliable estimates of marginal cost, the ultimate goal of the exercise)."[5]

Fundamentally, however, multicollinearity is not a problem. Numerous econometric textbooks attest to this fact.

---

1 Declaration of Christian T. Lundblad on Behalf of Amazon Fulfillment Services, Inc., Docket RM2015-7, July 8, 2015 ("Lundblad Report"), page 6.

2 Lundblad Report, page 7.

3 Lundblad Report, page 9.

4 Lundblad Report, page 17.

5 Lundblad Report, page 15.

> "So the problem of multicollinearity when estimating a conditional expectation function in a multivariate population is quite parallel to the problem of small sample size when estimating the expectation of a univariate population. But researchers faced with the latter problem do not usually dramatize the situation, as some appear to do when faced with multicollinearity."[6]
>
> "the OLS estimator in the presence of multicollinearity remains unbiased and, in fact, is still the best linear unbiased estimator."[7]
>
> "since multicollinearity violates none of our assumptions, the 'problem' of multicollinearity is not really well-defined."[8]

Real econometric problems arise in applied econometric analyses in situations that cause the chosen estimation procedure to return biased coefficient estimates. In such situations an analysis can simply return wrong answers, leading the analyst to incorrect conclusions about the nature and magnitudes of the effects he is trying to capture. Serious problems can also arise under conditions that cause the chosen estimation procedure to return biased estimates of the standard errors of coefficients, causing the analyst to misjudge the precision and reliability of his findings.

Multicollinearity causes no such problems. Assuming that other conditions are met, multiple regression analysis will return unbiased estimates of coefficient values and standard errors, even in the presence of multicollinearity. Despite the presence of multicollinearity, what you see is what you get.

Indeed, one can argue that one of the purposes of multiple regression analysis is to permit one to disentangle individual associations in the presence of collinear regressors. If all of the independent variables one wished to include in a regression analysis were perfectly orthogonal one could obtain unbiased estimates of their individual effects by including them in a series of univariate regression analyses. It is only when there is some degree of correlation among one's regressors (admittedly, a very, very common occurrence) that leaving important explanatory variables out of a model will lead to biased coefficient estimates – a truly serious econometric problem.

---

6 Arthur S. Goldberger, "A Course in Econometrics", page 246.

7 Peter Kennedy "A Guide to Econometrics", 6th Edition, page 193.

8 Jeffrey M. Wooldridge "Introductory Econometrics: A Modern Approach", 4th Edition, page 97.

This is not to say that multicollinearity is not an issue that an analyst should be aware of and take into account. As one's regressors become highly correlated the amount of "new" information that each adds to the analysis becomes smaller. This fact means that in a very real sense, as the degree of multicollinearity increases, one has less and less information to inform the computation of coefficient estimates. It is for this reason that Goldberger characterized multicollinearity as being "parallel to the problem of small sample size." In the presence of multicollinearity it takes larger sample sizes to achieve the same degree of statistical precision. However, in order to judge whether an adequate sample size has in fact been achieved, all one needs to do is examine the standard errors reported as part of the normal regression analysis results. They will tell an accurate story.

One fact worth noting in the context of this discussion is that the flexible functional form that the Postal Service has advocated and that the Commission has regarded with approval in other cost analyses is virtually guaranteed to produce a set of regressors characterized by a high degree of multicollinearity. Virtually any variable will wind up being highly correlated with its square, or with cross products between itself and other variables in the model. This fact does not mean that such a model is fatally flawed. It does mean, however, that one needs to make sure that one has access to a large enough sample to meet its more demanding estimation requirements – a point I made repeatedly in my initial report in this proceeding.

High degrees of multicollinearity can create a situation in which a linear combination of coefficients can be estimated with a high degree of precision even though the individual coefficients have high standard errors. Such a situation is akin to watching two people pushing a boulder up a hill. One might not be able to tell how hard either one is pushing (because they are pushing at the same time), but one can determine how hard the two of them are pushing together by measuring the progress of the boulder up the hill.

I mention this point because of its relevance to some of the comments of Professors Lundblad and Bradley that I discuss in more detail below. Many of the comments focus upon the imprecision with which individual coefficients are estimated. In weighing such comments it is important to keep in mind that neither the Commission, the Postal Service nor UPS should care whether, for example, the coefficient on the cross-product of DPS and FSS mail volumes is estimated with a high degree of precision. In the context of the flexible functional form that both I and the Postal Service have estimated, all that the Commission should really care about is the

precision with which marginal costs – computed as a linear combination of estimated coefficients – are estimated.

B. CRITICISMS OF THE NATIONAL FORM 3999 MODEL

Professor Lundblad argues that its reliance on imputed volume measures for certain mail streams renders the National Form 3999 Model useless for costing purposes. In the end, however, I don't believe he is able to provide the evidence needed to support this argument.

In addressing his comments, I must begin by noting that reliance on imputed values for certain mail streams cannot by itself serve as justification for rejection of the National Form 3999 Model. When I first suggested the possibility of using the National Form 3999 Data Set for the development of a city carrier costing model, it was clear that this data set lacked volume measures for certain mail streams. I suggested at the time that this deficiency could be cured by the development of appropriate imputation models.[9] Implicit in the Commission's order to the Postal Service to make the National Form 3999 Data Set available was an admission that my proposal offered a potentially acceptable solution. Otherwise there would have been no point in carrying out an analysis of the 3999 data. For this reason, I would argue, the National Form 3999 Model should not be rejected simply because it relies upon imputed regressors. Rather, the relevant question has to be whether the solution that I have devised is good enough to provide reliable results.

Professor Lundblad argues that the specific imputation models upon which the National Form 3999 Model relies are unacceptable because they have a "sizeable degree of multicollinearity" with the other volume variables included in the model.[10]

Professor Lundblad focuses in particular on my decision to include the number of delivery points in the zip code as an explanatory variable in the imputation models. He includes in his report a number of scatter plots showing the relationship between the number of delivery points in a zip code and the imputed values for deviation parcel, in-receptacle parcel and collection mail volumes. He argues that "the three imputed volumes (particularly the parcel volumes) that Dr. Neels employs in the second stage costing regression are largely determined by the same

---

9 Report of Kevin Neels on Behalf of United Parcel Service, Docket RM2015-7, March 18, 2015 ("Initial Neels Report"), pages 26-27.

10 Lundblad Report, paragraph 25.

explanatory variable (and one that is already an explanatory variable in its own right in the same regression)."[11] Finally, he tests a number of alternative specifications of the first stage imputation models, notes that the second stage results are sensitive to these specification changes, and argues as a result that "economic inference about relevant costs (is) impossible."[12]

I am not persuaded by Professor Lundblad's arguments on this point, and I don't believe that the Commission should be either.

First, as I have discussed above, multicollinearity is simply not a fatal flaw in a regression analysis. Some degree of collinearity is present in virtually all such analyses, and it poses a problem only to the extent that it prevents one from measuring the quantities of interest with acceptable precision.

Second, as regards the inclusion of delivery point counts in the imputation models, I would argue that the number of delivery points in a zip code is the one variable above all others that has to be included in any plausible mail volume model. From a mail volume generation standpoint the number of delivery points is the most fundamental measure of the "size" of a zip code. It is for this reason that all of the mail volume measures included in the model are correlated with the number of delivery points, as Table 1 below shows. The correlation between DPS mail volumes and the number of delivery points exceeds .9 in the Form 3999 Data Set. The correlations between cased mail volumes and sequenced mail volumes and the number of delivery points exceed .75 and .65 respectively. Only FSS mail volumes, which appear only in a limited set of zip codes, have a relatively low correlation with the number of delivery points. Far from being a fatal flaw, I would argue that the inclusion of delivery point counts in the imputation models is an essential prerequisite for their reliable performance.

---

11 Lundblad Report, pages 10-12.

12 Lundblad Report, paragraphs 35-39.

Table 1: Correlation between Mail Volumes and the Number of Delivery Points in National Form 3999 Data Set

Volume Variable	Correlation
3999 Volumes	
DPS Letters	0.9123
Cased Mail	0.7589
FSS Mail	0.3065
Sequenced Mail	0.6557
Imputations	
In-Receptacle Parcels	0.9527
Deviation Parcels	0.9938
Collection Volume	0.8674
Parcels (Deviation + In-receptacle)	0.9785

Finally, I question the conclusions Professor Lundblad draws from his tests of alternative imputation model specifications. The design of these tests virtually guarantees that they will fail. Starting with the specifications presented in my June Report, he strips away successive sets of explanatory variables, arriving at the end at specifications with half or less of the explanatory power of what he started with. It is not surprising that these stripped-down models do such a poor job of explaining variations in gross street time.

C. Professor Lundblad Does Not Address My Criticisms of the Postal Service's Special Study Panel Data Strategy

Although Professor Lundblad strongly defends the Postal Service's reliance on a panel data set in which the temporal dimension consists of two weeks of daily observations, he never comments upon my argument that much of this temporal variation consists of predictable day of the week related volume differences that would have been taken into account by the Postal Service in designing letter carrier routes.[13] More generally, he displays little awareness of the route restructuring process, or of the importance of accounting for the effects of volume change on route structure.

[13] Lundblad Report, paragraphs 52-55. Report of Kevin Neels on Behalf of United Parcel Service, Docket RM2015-7, June 8, 2015 ("Neels June Report"), page 8.

Unlike in financial markets, the primary focus of Professor Lundblad's research, which can respond nimbly and almost instantaneously to changes in market conditions, the operations and organization of the Postal Service change much more slowly. The Postal Service today will look an awful lot like it did yesterday. Over the course of a year, however, the ongoing route restructuring process undertaken by the Postal Service can bring about substantial changes in the structure and of the delivery process.

It is for this reason that I argued in my last two reports and continue to argue here that the panel dimension of the parcel and collection mail special studies adds data points to the analysis, but little useful information.[14] It might well be the case that on a low volume Friday a letter carrier might be able to complete his deliveries a little earlier than on a busy, high volume Monday. But the difference in street time between these two days will be small relative to what could be achieved if there were a sustained and across the board reduction in volume to which the Postal Service could respond by restructuring its routes and delivery operations to permit more efficient utilization of its human and material resources.

D. Professor Lundblad Does Not Address My Concerns Regarding Reliance on Special Field Studies

Professor Lundblad dismisses my concerns over the Postal Service's reliance on field studies to measure important costing parameters, arguing that there is a long history in economics, marketing, and related fields of using data collected through field studies.[15] While much of what he says here is true, it is also irrelevant to the basis for my concerns. These concerns stem from the fact that these studies are costly, and as a result, often underpowered, and so infrequently updated that their use forces the Commission to base important regulatory decisions on information and analyses that are often woefully out of date. I never argued that special studies represented an illegitimate approach, or one that was doomed to produce invalid results. Rather, the basis for my concerns was the fact that their cost would inevitably create tension between the limits of available financial resources, and both the sample sizes required to produce statistically reliable results and the necessity of carrying out repeated studies in order to keep up with the rapidly changing postal environment. None of Professor Lundblad's comments acknowledge or rebut these very real and practical concerns.

---

[14] Initial Neels Report, pages 14-17. Neels June Report, pages 7-11.

[15] Lundblad Report, page 24, paragraph 57.

IV. Comments on the Bradley Report

Like Professor Lundblad, Professor Bradley criticizes the use of imputed volume measures in the National Form 3999 Model.[16] To evaluate the accuracy of the models used to generate those imputed volumes, Professor Bradley conducts a test in which he estimates the deviation parcel volume equation using just the data from the first week of the parcel study, and then uses that model to forecast parcel volumes for the second week.[17] He concludes that the forecasting errors are "very large,"[18] although he does not posit any particular standard of comparison. I discuss these "out of sample" tests in more detail below.

Professor Bradley notes that the data used to estimate the volume imputation models came from a study conducted in 2014, while the data contained in the Form 3999 Data Set span a period of time reaching as far back as 2010. He speculates that, given the growth in parcel volumes that took place over this time period, there was some likelihood that the imputations were over-predicting parcel volumes for the earlier observations.[19] I address this point in more detail below as well.

Professor Bradley echoes Professor Lundblad in expressing concern over the problem of multicollinearity and over the possibility of data errors in the DOIS parcel volumes. He speculates that the imputed parcel variables may be picking up part of the influence of delivery point counts and DPS mail volumes on delivery times.[20] However, I don't see how this could happen if the delivery point and DPS mail volume variables are included in the 3999 model, as is in fact the case. Professor Bradley also raises concerns over the magnitudes of the marginal and incremental parcel delivery times implied by the National Form 3999 Model.[21]

Professor Bradley also offers a number of comments on the Modified Proposal Thirteen alternative discussed in my June Report. He argues that many of the activities in which a letter carrier is handling multiple mail streams are likely to have costs that are insensitive to changes in

---

16 Analysis of the Supplemental Report of Dr. Kevin Neels On Behalf of United Parcel Service, Docket RM2015-7, July 8, 2015 ("Bradley Report"), page 3.

17 Bradley Report, pages 5-10.

18 Bradley Report, page 10.

19 Bradley Report, page 10.

20 Bradley Report, page 13.

21 Bradley Report, pages 15-17.

volume, and that as a result there is no reason to consider including parcel volumes in a regular delivery time equation.[22] He also presents the results of a number of statistical tests, which, he argues, support the exclusion of parcels volumes from a regular delivery time model.[23] I do not find either his argument or his statistical tests to be persuasive, and I discuss them in more detail below.

V. Updates to the National Form 3999 Model

In an effort to respond to some of the comments offered by Professors Bradley and Lundblad on the National Form 3999 Model presented in my June Report, I have carried out a number of changes to this analysis. These changes are designed to address specific points raised in their reports. First, Professor Bradley has argued that the imputation models presented in my June Report fail to account for changes in parcel volumes that have occurred in recent years, and the implications of those changes. He questioned the ability of regression equations estimated on 2014 special study data to accurately predict parcel volumes for the earlier dates contained within the Form 3999 Data Set.[24] Second, Professor Bradley has raised questions about the out of sample forecasting properties on the imputation models presented in my June Report.[25] Finally, both Professors Lundblad and Bradley have questioned the reliability of the DOIS parcel volume data used in those imputation models.[26]

In addressing these concerns I have attempted to stay as close as possible to both the general spirit and the specific features and details of the analysis presented in the June Report. The specific details of how I addressed these concerns are discussed below.

A. Imputation Model Changes

I made two changes to the first stage volume imputation models that I believe have significantly improved the reliability of the National Form 3999 Model. The first of these involved substitution of simple linear regression models for the negative binomial models included in my

---

22 Bradley Report, pages 19-20.

23 Bradley Report, pages 20-24.

24 Bradley Report, page 10.

25 Bradley Report, pages 5-10.

26 Lundblad Report, pages 19-20. Bradley Report, page 23, footnote 22.

June Report. The second involved the incorporation of a series of volume indexes to improve the ability of the models to track aggregate changes in volumes over time.

My original decision to estimate volume imputation models using a negative binomial model was motivated by the fact that this model is designed to handle count data in which the dependent variable of the model takes the form of integers. The negative binomial model also generated non-negative (though not necessarily integer) predictions. Professor Bradley in his report presented the results of a simple out of sample forecasting test in which he re-estimated the imputation models from my June Report using the first week of parcel and collection volume study data, and then evaluated the ability of those models to predict the actual second week values.[27] He reported some discrepancies between predicted and actual values, and argued that "These results raise serious questions about the utility of the imputations in estimating the street time variability equation."[28] In the course of investigating Professor Bradley's research on this point I determined that the simple linear regression models appeared to have better out of sample forecasting properties, and seemed to mitigate somewhat the forecasting problems he identified. Compared to this advantage, the benefits of using a model suited to integer data appeared trivial.

Table 2 and Table 3 compare the out of sample forecasting abilities of the negative binomial and linear imputation models. Table 2 presents the results of the Mincer-Zarnowitz test, which Professor Bradley uses in his report, and which is based upon the results of regressing the out of sample values on the out of sample model predictions. For all three volume variables the linear model slope coefficients are closer to one and the constants are closer to zero than for the negative binomial model. Table 3 employs another test used by Professor Bradley, and compares the mean and median absolute and percentage errors of the two models. For comparison purposes I have also included in these tables measures of forecasting accuracy for a set of "best case" forecasting models. These models take the actual parcel and collection mail volumes for a specific day of the first week of the study as the prediction for the corresponding day of the second week.[29]

---

[27] Bradley Report, pages 5-10.

[28] Bradley Report, page 10.

[29] As I describe below, in the imputation strategy used in this update I employ a set of purely cross-sectional imputation models. Given this choice, the perfect forecasting model for Professor Bradley's test would be one that predicts the first week volumes with 100 percent accuracy.

It is interesting to note that by the Mincer-Zarnowitz test both the negative binomial regression and linear regression models outperform the first week test. This fact suggests to me that the regression based imputation models may actually do a superior job of separating the systematic determinants of these volume measures from random fluctuations. However, the first week test results in substantially lower mean absolute and percentage errors. The superior forecasting ability of the naïve first week model, in combination with the results of the Mincer-Zarnowitz test, suggests that there may be random factors influencing these volume variables that persist over time. Such a finding would be consistent with a point I have long made – namely, that adding additional consecutive days to a panel study of delivery costs adds little additional information.

The forecasting performance of the imputation models should be judged relative to the performance of the naïve first week model. This comparison suggests that for the parcel variables the imputation models may be doing a fairly reasonable job, but that the imputation model for collection volume seems to be performing less well.

For the two parcel equations the prediction errors of the linear parcel equations are smaller than those of the negative binomial models. For the collection mail equation the prediction errors of the linear version are somewhat larger than those of the negative binomial equation, even though the negative binomial performs less well on the Mincer-Zarnowitz test. Overall, from a forecast accuracy standpoint the linear versions appeared superior.

Table 2: Mincer-Zarnowitz Test on Negative Binomial and Linear Models

Model	Coefficient	Estimated Value	T-Stat	95% Confidence Interval (Low, High)
In Receptacle Parcels				
[A] Negative Binomial	β_0	13.310	1.26	(-7.450, 34.080)
	β_1	0.935	50.90	(0.899, 0.971)
[B] Linear	β_0	-4.627	-0.44	(-25.470, 16.210)
	β_1	0.971	52.31	(0.934, 1.007)
[C] First Week	β_0	55.080	7.47	(40.620, 69.540)
	β_1	0.863	70.62	(0.839, 0.887)
Deviation Parcels				
[A] Negative Binomial	β_0	21.840	2.65	(5.655, 38.030)
	β_1	0.900	42.51	(0.858, 0.941)
[B] Linear	β_0	-0.390	-0.05	(-16.090, 15.310)
	β_1	0.971	46.71	(0.930, 1.011)
[C] First Week	β_0	35.840	7.40	(26.340, 45.330)
	β_1	0.879	74.33	(0.856, 0.902)
Collection Volume				
[A] Negative Binomial	β_0	-274.000	-1.77	(-578.300, 30.210)
	β_1	1.086	27.50	(1.009, 1.164)
[B] Linear	β_0	103.600	0.72	(-178.500, 385.800)
	β_1	0.984	27.27	(0.913, 1.055)
[C] First Week	β_0	316.100	6.38	(218.900, 413.200)
	β_1	0.929	89.21	(0.908, 0.949)

Table 3: Prediction Error for Negative Binomial and Linear Models

Model	Mean Absolute Prediction Error	Mean Absolute Percentage Prediction Error	Median Absolute Prediction Error	Median Absolute Percentage Prediction Error
In Receptacle Parcels				
[A] Negative Binomial	146.65	71.90%	103.14	26.00%
[B] Linear	143.18	68.52%	100.12	24.34%
[C] First Week	95.94	29.20%	56.00	14.43%
Deviation Parcels				
[A] Negative Binomial	114.62	88.22%	80.14	29.45%
[B] Linear	107.29	71.86%	73.78	27.17%
[C] First Week	67.48	29.59%	41.00	16.13%
Collection Volume				
[A] Negative Binomial	1,968.14	262.02%	1,440.34	50.30%
[B] Linear	2,003.59	293.91%	1,528.66	50.85%
[C] First Week	874.10	45.65%	490.83	20.62%

Notes:

[A]: Negative binomial model as specified in June Report.

[B]: Linear version of [A].

[C]: Treats first week reported volumes as second week predictions for the same zip code and day of the week.

These findings convinced me that the switch to simple linear regression based imputation models would provide more robust and reliable estimates of street time variabilities.

My decision to incorporate intertemporal volume change indexes into the imputation models was based on a thoughtful observation made by Professor Bradley. He noted in his report that while the Form 3999 data span a multiyear period from 2010 through 2013, the parcel study data on which my imputation models were based are drawn entirely from a two month period in 2014. He questioned whether, given the growth in parcel volumes that has taken place in recent years, a model based on 2014 data would be capable of producing accurate parcel volume

estimates for earlier years.[30] This observation seemed to raise a reasonable point, and one that needed to be addressed.[31]

The solution I devised was to remove from the imputation models all of the time varying mail volume variables, and then to combine the forecasts produced by these models with a series of mail stream specific indexes measuring changes over time in the volumes of the specific products comprising each mail stream. This solution turned the computation of imputed volumes effectively into a two-step process. For each zip code in the Form 3999 Data Set I first calculated what was effectively a set of March/April 2014 estimates of deviation parcels and in-receptacle parcels and a set of April/May 2013 collection mail volumes, and then adjusted each of these estimates based on mail stream specific volume changes between the special study time periods and the "as of" date of the information contained in the Form 3999 Data Set for that zip code.[32]

The "stripped down" imputation regression models for deviation parcels, in-receptacle parcels and collection volumes are shown in Table 4, Table 5, and Table 6.

---

30 Bradley Report, page 10.

31 This same point had occurred independently to me and the members of my team, and so we have been considering potential ways of addressing it.

32 This procedure assumes implicitly that over the period covered by the Form 3999 data the three imputed mail streams have been growing at the same rate in all zip codes. While I doubt that this assumption is *precisely* true, I would not expect departures from strict proportionality to introduce any specific bias into the results. I do believe that this a priori assumption produces more appropriate estimates of parcel and collection mail volumes for the earlier years. In any case, it is the only assumption we can make, given the absence of deviation parcel, in-receptacle parcel and collection volume data for any other point in time.

Table 4: Linear Deviation Parcels Imputation Regression

Variable	Estimated Coefficient	Cluster-Robust T-Statistics
Delivery Points (DP)	0.0256	22.45
DOW=Tuesday	(62.6)	(11.31)
DOW=Wednesday	(8.37)	(2.22)
DOW=Thursday	2.36	0.62
DOW=Friday	(2.32)	(0.65)
DOW=Saturday	(19.2)	(5.05)
R^2	0.613	
Observations	3,321	

Table 5: Linear In-Receptacle Parcels Imputation Regression

Variable	Estimated Coefficient	Cluster-Robust T-Statistics
Delivery Points (DP)	0.0386	21.19
Population per DP	1.75	0.64
Employment per DP	(2.78)	(0.50)
Mean Household Income	0.00219	8.55
NAICS 11: Agriculture, Forestry, Fishing and Hunting	(294)	(0.24)
NAICS 21: Mining, Quarrying, and Oil and Gas Extraction	(336)	(1.06)
NAICS 22: Utilities	(146)	(0.25)
NAICS 23: Construction	2.46	0.01
NAICS 31: Manufacturing	(166)	(0.85)
NAICS 42: Wholesale Trade	100	0.49
NAICS 44: Retail Trade	(40.8)	(0.21)
NAICS 48: Transportation and Warehousing	(404)	(1.55)
NAICS 51: Information	(358)	(0.65)
NAICS 52: Finance and Insurance	(336)	(1.73)
NAICS 53: Real Estate and Rental and Leasing	(991)	(2.02)
NAICS 54: Professional, Scientific, and Technical Services	(199)	(1.00)
NAICS 55: Management of Companies and Enterprises	(516)	(1.86)
NAICS 56: Administrative and Support	(31.1)	(0.13)
NAICS 61: Educational Services	(545)	(2.72)
NAICS 62: Health Care and Social Assistance	(161)	(0.81)
NAICS 71: Arts, Entertainment, and Recreation	194	0.69
NAICS 72: Accommodation and Food Services	(282)	(1.29)
NAICS 99: Unclassified	(8,553)	(0.73)
DOW=Tuesday	(170)	(14.77)
DOW=Wednesday	(84.7)	(11.45)
DOW=Thursday	(59.4)	(8.92)
DOW=Friday	(54.4)	(9.03)
DOW=Saturday	(75.1)	(10.96)
R^2	0.725	
Observations	3,319	

Table 6: Linear Collection Volume Imputation Regression

Variable	Estimated Coefficient	Cluster-Robust T-Statistics
Delivery Points (DP)	0.213	8.73
NAICS 56: Administrative and Support	10,104	3.12
DOW=Tuesday	(161)	(3.76)
DOW=Wednesday	(35.2)	(0.74)
DOW=Thursday	(14.2)	(0.30)
DOW=Friday	(107)	(2.10)
DOW=Saturday	(1,330)	(12.15)
Age Group: Under 10	(58.21)	(1.60)
Age Group: Between 10 and 19	(9.80)	(0.30)
Age Group: Between 30 and 39	(4.82)	(0.08)
Age Group: Between 40 and 49	50.3	1.15
Age Group: Between 50 and 59	(2.29)	(0.04)
Age Group: Between 60 and 69	(35.4)	(0.63)
Age Group: Between 70 and 79	(92.8)	(1.05)
Age Group: Over 80	195	2.16
Census Division 2: Middle Atlantic	242	0.49
Census Division 3: East North Central	575	1.12
Census Division 4: West North Central	1,459	1.94
Census Division 5: South Atlantic	(82.1)	(0.16)
Census Division 6: East South Central	361	0.65
Census Division 7: West South Central	(46.8)	(0.08)
Census Division 8: Mountain	1,264	1.93
Census Division 9: Pacific	(247)	(0.48)
R^2	0.387	
Observations	3,498	

To construct the volume indexes associated with these three mail streams I combined information from the distribution keys for these three mail streams from RM 2015.7.1 with annual piece counts from the annual Revenue, Pieces and Weights (RPW) reports and select information from ACR2010 RPW by Shape and Indicia. The distribution keys describe the composition of each mail stream by product. I converted the distribution keys into percentages by product, and the piece counts from the RPW reports into two sets of volume indexes – one that took values of one in March/April of 2014 (when the parcel special study data were

collected), and a second that took values of one in April/May of 2013 (when the collection volume special study data were collected). Using the distribution key percentages as weights, I then constructed a volume index for each of the three imputed mail streams. These volume indexes are shown in Table 7. They do show the increase in parcel volumes in recent years. They also show that the products that make up collection mail have declined sharply over this same period.

Table 7: Imputed Variable Volume Trends

Year	In-Receptacle Parcels	Deviation Parcels	Collection Volume
2008	0.844	0.798	1.538
2009	0.768	0.722	1.366
2010	0.764	0.717	1.242
2011	0.816	0.754	1.137
2012	0.884	0.822	1.054
2013	0.920	0.926	1.000
2014	1.000	1.000	0.960

Note: Parcel trends are scaled to 1 in 2014.
Collection volume trends are scaled to 1 in 2013.

Combining the regression results shown in Table 4, Table 5, and Table 6 and the volume indexes shown in Table 7, I computed the imputed volume measures used in the National Form 3999 Model as:

$$\widehat{y_i} = \tilde{y}_i I_{t(i)}$$

Where $\tilde{y}_i$ is the value predicted by the imputation regression model for zip code *i*, *t(i)* is the time period to which the Form 3999 data for zip code *i* refers, and I_t is the volume index value for period *t*. In order to account more accurately for when within the year the Form 3999 data were collected, I converted the annual volume indexes shown in Table 7 into monthly values by linear interpolation.

To verify the accuracy and reliability of these imputations I conducted two tests that I believe are both more rigorous and more relevant than the first week/second week test conducted by Professor Bradley.

The Form 3999 Data Set contains a field labeled "Parcel Hours" which, according to the Postal Service, measures "all instances when additional time over loading mail into the receptacle was needed for handling any parcel, regardless of its size/weight."[33] This definition implies that this field measures "deviation time" – extra time required for parcel delivery, over and above what would be required simply to place the parcel in the mail receptacle. Summing this quantity across the entire Form 3999 dataset and dividing by the associated imputed parcel volumes whose derivation is described immediately above yields a value of 41.3 seconds per piece. To place this quantity in perspective, I note that the average scan time per deviation or accountable piece reported by the parcel study was 52.6 seconds per piece. The corresponding figure for in-receptacle parcels was 28.5 seconds per piece. Together these figures imply that the per piece amount of "additional time over loading mail into the receptacle" was 24.1 seconds per piece – a figure somewhat below the 41.3 second calculated using the Form 3999 data, but still remarkably close given the challenges of imputing volume measures across more than 10,000 zip codes and the fact that the times in question come from completely independent datasets constructed for different purposes during time periods many years apart. The tight correspondence between these two figures indicates that for deviation parcels, at least, the imputed volume levels are very close to where they should be.

From these results alone one cannot exclude the possibility that despite the fact that parcel volumes sum to more or less the correct total, they might still be unusably noisy at the individual zip code level. To exclude that possibility I conducted another test that involved regressing the log of DOIS parcel volumes from the Form 3999 Data Set on the log of the sum of the imputed values for deviation parcels and in-receptacle parcels.[34] Results of this regression are shown below in Table 8. The log of the sum of the imputed parcel volumes by itself explains over 98 percent of the variation in the log of DOIS parcel volumes. In connection with this result it is worth noting that the imputation models used to construct the two parcel variables included in the single right hand side variable did not include any time varying mail volume measures. These results indicate that the imputed parcel volume variables not only sum to the correct overall volume level, but are also distributed across zip codes in a way that tracks very closely a related volume measure that is collected completely independently.

---

[33] NOTICE OF THE UNITED STATES POSTAL SERVICE OF PROVIDING INFORMAL RESPONSES TO UPS QUESTIONS, May 28, 2015, Response to Question 3 (a).

[34] The Postal Service indicated that the DOIS parcel counts could include either deviation parcels of in-receptacle parcels. See NOTICE OF THE UNITED STATES POSTAL SERVICE OF PROVIDING INFORMAL RESPONSES TO UPS QUESTIONS, May 28, 2015, Response to Question 1 (a).

Table 8: Prediction Power of Imputed Parcels

Variable	Coefficient	T-Stat
Log(Imputed Parcels Variable)	0.839	725.2
R^2	0.981	
Observations	10,049	

B. UPDATED NATIONAL FORM 3999 MODEL RESULTS

I re-estimated the National Form 3999 Model, making two changes that I believe have greatly increased its reliability. The first involved the use of the improved imputation models described above. The second was to sum the imputed deviation parcel and in-receptacle parcel volumes, and to enter the result into the model as a single composite parcel volume measure. My motivations for making the latter change were partly statistical and partly operational.

The idea of combining these two variables initially arose as a possible way of dealing with statistical problems. I had noticed over the course of working with the model a tendency for the estimated coefficients for the two parcel variables to shift in interrelated ways in response to small changes in model estimation or specification. This type of behavior is known to occur when two regressors are highly correlated. As I noted above, in such situation it is often possible to estimate the coefficient of a linear combination of two such regressors (e.g., a sum) much more precisely than the individual coefficients can be estimated. These facts led me to consider whether I might derive more reliable results by combining them.

In considering whether such a change made operational sense I reviewed some of the responses of the Postal Service to questions regarding the definitions of the various parcel variables, and their relationships to other variables in the Form 3999 Data Set. In these responses, the Postal Service stated that the Form 3999 Data Set recorded some positive amount of "parcel time" when "additional time over loading mail into the receptacle was needed for handling any parcel, regardless of its size/weight."[35] In that same response the Postal Service noted that a small parcel

[35] NOTICE OF THE UNITED STATES POSTAL SERVICE OF PROVIDING INFORMAL RESPONSES TO UPS QUESTIONS, May 28, 2015, Response to Question 3 (a).

might require additional handling time "because it does not fit in the receptacle."[36] These statements indicate that a letter carrier will not know whether a particular parcel will turn out to be a deviation parcel until he or she reaches the mail receptacle. This determination will depend upon the relationship between the size of the parcel and the size of the receptacle, perhaps also on the amount of other types of mail that might be present, and whether that mail exhausts the volume of the receptacle. These facts indicated to me that deviation parcels and in-receptacle parcels were essentially different components of the same mail stream.

In my June Report I had recommended that the Commission direct the Postal Service to start collecting deviation parcel, in-receptacle parcel and collection mail volume data as part of the route evaluation process that generates the information included in the Form 3999 Data Set. In considering this recommendation in light of the facts outlined above, I realized that the collection of volume counts for deviation and in-receptacle parcels would require the letter carriers or their supervisors to maintain running counts of which category each parcel fell into as it was delivered. In contrast, a single undifferentiated count of the two categories of parcels could be carried out before the letter carrier left the delivery unit. I suspected that the reason why the DOIS parcel counts are based upon a size and weight definition that does not reflect whether a parcel requires additional handling time was to permit these counts to be carried out at the delivery unit, rather than along the route.[37] These considerations led me to believe that dropping the distinction for modeling purposes between deviation and in-receptacle parcels would both be consistent with operational realities and make it easier for the Postal Service to comply with a directive to begin collecting accurate parcel count data as part of its route evaluation process.

I recognize that there are cost consequences between deviation and in-receptacle parcels that need to be taken into account. However, there are ways of accomplishing this that do not require that these two categories of parcels be entered into a street time econometric model as separate variables. Another statistical system currently maintained by the Postal Service – the City Carrier Costing System – already collects city carrier mail volume and characteristics data for a

---

[36] NOTICE OF THE UNITED STATES POSTAL SERVICE OF PROVIDING INFORMAL RESPONSES TO UPS QUESTIONS, May 28, 2015, Response to Question 3 (c).

[37] DOIS parcels include parcels that meet either of the following two criteria: 1) larger than a shoebox or 2) heavier than two pounds. NOTICE OF THE UNITED STATES POSTAL SERVICE OF PROVIDING INFORMAL RESPONSES TO UPS QUESTIONS, May 28, 2015, Response to Question 1 (a).

representative national sample of routes.[38] These data can be used to calculate how the parcel mail stream is divided between deviation and in-receptacle parcels, and to track changes in these proportions over time.

All of these factors led me to the conclusion that combining the two parcel categories into a single volume measure was an appropriate change to make in the National Form 3999 Model.

The updated National Form 3999 Model regression results are shown in Table 9, alongside the comparable results from my June Report. The combined parcel variable in the updated model is associated with a marginal cost and variability that falls between the previously estimated variabilities for deviation and in-receptacle parcels, but is much closer to the previous in-receptacle parcel value. Estimates of marginal cost and variability for the other mail streams are broadly similar to the previously reported estimates, although slightly higher.

It is important to point out that these variabilities are illustrative only, and should not be used directly for costing purposes. As shown in the December 2014 City Carrier Street Time Study, the formula for calculating the volume variability for a particular product requires data on the volume for this product and for the total delivery time in the target year.[39] The variabilities and marginal costs shown in Table 9 are based upon the averages of the volumes and times from the 3999 dataset, which refer to no single point in time. To apply these results to calculate product level costs one would need to combine the coefficients from the regressions shown in Table 9 with volume and time data for the target year.

---

[38] Scoping Study Report of the United States Postal Service, Priorities for Future Data Collection and Analytical Work Relating to Periodic Reporting, Docket RM2011-3, May 25, 2012, pages 13-14.

[39] Report on the City Carrier Street Time Study, Docket RM2015-7, December 2014, page 67.

Table 9: 3999 Model Results Comparison

	3999 Model from June 2015		July 2015 Update: Linear Imputations with Volume Trend	
Variable	Coefficient	Robust standard error	Coefficient	Robust standard error
	[1]	[2]	[3]	[4]
dps_adj	-2.50E-04	1.99E-04	2.29E-04	1.95E-04
dps_adj_2	-3.24E-08	8.97E-09	-2.41E-08	8.14E-09
dps_fss_adj	9.23E-08	3.20E-08	8.80E-08	3.05E-08
dps_seq_adj	-1.53E-08	2.11E-08	-4.52E-08	2.08E-08
dps_cm_adj	4.91E-08	1.98E-08	6.38E-08	1.90E-08
dps_dev_par	2.17E-06	1.47E-06		
dps_cv	9.47E-08	5.33E-08	1.51E-07	7.34E-08
dps_par			-9.52E-08	8.47E-07
dps_irp	-1.39E-07	6.87E-07		
dps_dp	8.98E-08	2.71E-08	8.19E-08	4.68E-08
dps_dm	1.56E-03	1.65E-04	1.28E-03	1.58E-04
dps_smdp	-2.02E-03	7.54E-04	-2.75E-03	8.59E-04
fss_adj	3.56E-03	4.69E-04	4.06E-03	5.15E-04
fss_adj_2	-1.11E-07	5.08E-08	-1.17E-07	5.30E-08
fss_seq_adj	2.61E-08	4.32E-08	4.07E-08	4.52E-08
fss_cm_adj	-5.52E-08	4.40E-08	-2.06E-08	4.48E-08
fss_dev_par	-2.73E-06	2.93E-06		
fss_cv	-3.05E-07	1.13E-07	-5.04E-07	1.72E-07
fss_par			-7.26E-07	1.91E-06
fss_irp	3.43E-06	1.45E-06		
fss_dp	-2.00E-07	6.11E-08	-6.79E-08	1.09E-07
fss_dm	-6.59E-04	3.72E-04	-8.75E-04	3.81E-04
fss_smdp	1.03E-02	8.64E-03	1.05E-02	8.79E-03
seq_adj	5.42E-04	3.05E-04	-7.31E-05	3.25E-04
seq_adj_2	-1.51E-07	2.72E-08	-1.71E-07	2.78E-08
seq_cm_adj	9.07E-08	3.02E-08	1.18E-07	3.20E-08
seq_dev_par	-5.13E-06	2.90E-06		
seq_cv	-3.33E-07	8.35E-08	-3.24E-07	1.19E-07
seq_par			1.99E-06	1.64E-06
seq_irp	2.23E-06	1.34E-06		
seq_dp	2.23E-07	5.03E-08	1.75E-07	8.83E-08
seq_dm	2.77E-04	2.73E-04	5.13E-04	2.87E-04
seq_smdp	-1.57E-03	1.19E-03	-2.14E-03	1.02E-03
cm_adj	2.00E-03	3.05E-04	2.59E-03	3.15E-04
cm_adj_2	-3.04E-08	1.49E-08	-3.07E-08	1.63E-08
cm_dev_par	1.50E-06	1.84E-06		
cm_cv	-7.60E-08	7.75E-08	-1.41E-07	1.18E-07
cm_par			-1.75E-06	1.13E-06
cm_irp	8.23E-07	9.27E-07		
cm_dp	-1.80E-07	3.39E-08	-6.54E-08	6.48E-08
cm_dm	-1.38E-03	2.46E-04	-1.56E-03	2.53E-04
cm_smdp	1.53E-03	1.12E-03	1.83E-03	1.13E-03
cv	1.72E-04	6.84E-04	-1.57E-03	5.45E-04
cv_2	-2.25E-07	8.48E-08	2.72E-08	1.19E-07
cv_par			-5.92E-06	2.79E-06
cv_irp	-1.16E-06	2.63E-06		
cv_dp	5.26E-07	8.64E-08	5.80E-07	1.64E-07
cv_dm	1.29E-03	4.40E-04	1.03E-03	4.00E-04
cv_smdp	-6.48E-04	9.12E-04	7.61E-04	6.87E-04

Variable	3999 Model from June 2015 Coefficient [1]	3999 Model from June 2015 Robust standard error [2]	July 2015 Update: Linear Imputations with Volume Trend Coefficient [3]	July 2015 Update: Linear Imputations with Volume Trend Robust standard error [4]
dev_par	1.23E-01	2.11E-02		
dev_par_2	-2.03E-04	6.92E-05		
dev_par_cv	-8.05E-06	5.19E-06		
dev_par_irp	1.12E-04	7.20E-05		
dev_par_dp	3.83E-06	2.80E-06		
dev_par_dm	-9.72E-02	1.74E-02		
dev_par_smdp	-7.30E-02	4.00E-02		
par			6.51E-03	7.20E-03
par_2			3.11E-05	2.44E-05
par_dp			7.36E-07	2.13E-06
par_dm			-5.54E-03	6.05E-03
par_smdp			3.11E-03	1.53E-02
irp	3.89E-02	8.73E-03		
irp_2	-9.78E-05	2.30E-05		
irp_dp	6.86E-07	1.16E-06		
irp_dm	3.59E-03	7.79E-03		
irp_smdp	-3.25E-02	1.43E-02		
dp	4.79E-03	3.95E-04	6.41E-03	4.19E-04
dp_2	-2.73E-07	4.69E-08	-3.60E-07	7.79E-08
dp_dm	1.27E-03	2.54E-04	4.13E-04	3.32E-04
dp_smdp	3.36E-03	1.69E-03	2.48E-03	1.46E-03
dm	2.14E+01	2.21E+00	1.92E+01	1.74E+00
dm_2	-1.40E+01	1.48E+00	-1.63E+01	1.51E+00
dm_smdp	-1.00E-01	1.60E+00	-2.88E+00	1.41E+00
smdp	1.55E+01	2.81E+00	4.09E+00	1.24E+00
smdp_2	-3.71E-02	8.86E-03	-4.45E-03	5.09E-03
Constant	-1.61E+01	1.92E+00	-3.59E+00	6.54E-01
MARGINAL COSTS				
DPS Letters	3.60	0.29	4.09	0.29
FSS Mail	12.99	3.09	13.92	3.18
Sequenced Mail	1.41	0.42	1.51	0.38
Cased Mail	3.89	0.42	4.63	0.43
Collection Volume	5.74	0.65	6.22	0.77
Deviation Parcels	162.08	34.25		
In-Receptacle Parcels	38.42	10.49		
Parcels			52.43	10.36
VARIABILITIES				
DPS Letters	21.37%		24.77%	
FSS Mail	3.83%		4.19%	
Sequenced Mail	1.24%		1.36%	
Cased Mail	7.80%		9.47%	
Collection Volume	5.38%		5.66%	
Deviation Parcels	11.84%			
In-Receptacle Parcels	4.14%			
Parcels			8.03%	
Total	55.61%		53.46%	

Professor Bradley argued that some of the marginal cost estimates presented in the June Report were operationally unreasonable. Since he has raised this issue, I have attempted to compare these marginal costs against other available benchmarks. The most obvious such benchmarks are the scan times generated by the special parcel study. In that study letter carriers were asked to measure the direct time spent delivering in-receptacle parcels, deviation parcels, and accountables.

The instructions provided to letter carriers directed them to "start the clock" when initiating a delivery activity, and to "stop the clock" when the delivery was completed.[40] Based upon an econometric analysis of the data collected in this way the Postal Service concluded that the volume variability of in-receptacle parcels was 48.8 percent, and the combined volume variability of deviation parcels and accountables was 49.1 percent.[41]

I will admit to having a certain degree of skepticism regarding these volume variability estimates. Based simply upon the nature of the instructions provided to the letter carriers participating in the study, I am unable to envision a plausible situation in which elimination of the parcel would not generally result in the complete elimination of all of this scan time. The only plausible situation in which the reported scan time might NOT be entirely variable would seem to be in cases in which more than one in-receptacle parcel or deviation parcel/accountable were being delivered to the same delivery point on the same day. In such a situation one might plausibly anticipate that delivery of two pieces might take less than twice the time required to deliver one piece. However, given the number of parcels delivered on any given day and the number of delivery points over which they are spread, one would expect such multiple piece delivery events to be extremely rare unless it were the case that parcel deliveries were highly concentrated on a small number of delivery points.[42] In the absence of any evidence of such a highly concentrated pattern of deliveries, I am forced to conclude that the Postal Service's finding of low volume variability is likely some sort of statistical artifact.

One does not need to resolve this problem, however, in order to interpret the relative scan times produced by the parcel study as a measure of the relative work intensity of deviation and in-receptacle parcels. If I am correct in my suspicion that the reported scan times are entirely volume variable, one can interpret the relative scan times as a measure of the relative direct work content of the two mail streams. If, however, the Postal Service is correct in its assertion that the scans contain large amounts of fixed cost, one would have to scale the average per piece scan times down to arrive at a measure of the relative direct work content. Because the volume

---

40 Package_Accountable_Study_Exhibit_1- Carrier Study Guide.pdf, page 4.

41 Report on the City Carrier Street Time Study, December 2014, page 118. The value called out for deviation parcels and accountables is the sum of the reported variabilities for these two mail streams.

42 Based on the data collected in the special parcel study, the probability of an average delivery point receiving a deviation parcel on any given day is only 2.7 percent. This figure implies that the probability of any given delivery point receiving two parcels on any given day is only 0.07 percent.

variabilities for the two mail streams are essentially identical, however, such a calculation would take one to exactly the same place as my more skeptical interpretation.

Given a measures of relative work content and relative volumes for the deviation and in-receptacle parcels,[43] one can easily decompose the variable cost associated with the composite parcel variable into separate measures of the marginal time cost of deviation and in-receptacle parcels. I carry out this decomposition in Table 10. The deviation parcel scan time per piece is 53.6 seconds, and total marginal time per piece in 74.1 seconds. These figures imply (given my interpretation of the scan time results) that the additional time associated with whatever "friction" the presence of deviation parcels introduces in other letter carrier street activities amounts to 21.5 seconds per piece – a value that seems entirely reasonable. For in-receptacle parcels the direct handling time amounts to 28.5 seconds per piece, and the additional "friction" time to 11.6 seconds per piece, a result that also seems quite reasonable. These results strengthen my confidence that the National Form 3999 Model is producing accurate results.

Table 10: Estimation of In-Receptacle and Deviation Parcel Marginal Costs and Variabilities

		Combined Parcel Variable	
Variability	[A]	8.03%	
Marginal Cost (seconds)	[B]	52.43	
		Deviation Parcels [1]	In- Receptacle Parcels [2]
Average Scan Time	[C]	52.64	28.49
City Carrier DK Volumes	[D]	654,503	1,151,896
Total Estimated Scan Time	[E]	34,453,950	32,813,254
Scan Time Ratio	[F]	1.8	
Est. Variability	[G]	4.11%	3.92%
Est. Marginal Cost (seconds)	[H]	74.11	40.11

Notes and Sources:

[A], [B]: Variability and Marginal Cost output from Updated 3999 Model Results.

[C]: Average Scan Times from Deviation and In- Receptacle Special Studies.

[43] To measure relative volumes I rely in the distribution key information for these two mail streams, which is drawn from the City Carrier Costing System.

[C][1]: Deviation average scan time is total deviation time divided by accountable and deviation volume.

[D]: Total City Carrier Distribution Key Volumes for Large Parcels and Small Parcels.

[E]: [C] x [D]. This estimates the relative cost intensity of Deviation and In-Receptacle Parcels.

[F]: [C][1] / [C][2].

[G]: [A] weighted by [E].

[H][1]: [H][2] x [F].

[H][2]: [B] x sum([D]) / ([D][2] + [F] x [D][1]).

VI. Criticisms of Modified Proposal Thirteen

Both Professor Lundblad and Professor Bradley criticize what I have called the Modified Proposal Thirteen. I first summarize their criticisms; and then respond in detail to the more substantive points they raise.

A. Summary of the Criticisms Offered by Professor Lundblad

Professor Lundblad criticizes the Modified Proposal Thirteen model arguing that the introduction of additional parcel related variables creates a multicollinearity problem. He argues that "[t]he size of the VIFs for the parcel variables indicates that the coefficients on the parcel volume variables are picking up correlations to other variables, not a real relationship between parcel volume and regular delivery time."[44]

As I have discussed earlier in this report, multicollinearity is fundamentally not a problem in that it does not prevent the basic linear regression model from generating the best linear unbiased estimates. Respectfully, I do not see how Professor Lundblad's assertion can possibly be true. The VIF measure the extent to which one specific independent variable is correlated with the remaining variables in the model. If one were to take all of the remaining variables out of the model, it is true that a remaining collinear regressor would pick up part of their relationships to the dependent variable. But, this would not occur if all of the other regressors remained in the model. Professor Lundblad's argument that multicollinearity biases the relationship between parcel volume and regular delivery time is simply not true.

Professor Lundblad also believes the problems infecting the Form 3999 parcel volume affect the Modified Proposal Thirteen model. He seems to be primarily concerned about erroneous zero

[44] Lundblad Report, page 18.

values in the DOIS parcel data, but he also raises the possibility that other non-specified problems may be present. Professor Lundblad bases his conclusions on regressions of the deviation parcel volume on DOIS parcel volume per delivery point, and on DOIS parcel volume. He noted that these regressions had fairly low explanatory power, and took this as an indication of the presence of other errors in the DOIS parcel data. In discussing these results he does not mention (and may not be aware of) the fact that there are significant definitional differences between DOIS parcels on the one hand, and deviation and in-receptacle parcels on the other.[45]

B. SUMMARY OF THE CRITICISMS OFFERED BY PROFESSOR BRADLEY

Professor Bradley criticizes the Modified Proposal Thirteen on both a statistical and conceptual basis. His conceptual criticisms reflect a belief that the costs of activities involving mixed mail are fixed. I will address this point in more detail below.

1. Bradley Applies Inappropriate Statistical Tests

In evaluating whether the parcel-related variables belong in the regular delivery equation Professor Bradley fails to apply the standard and well-accepted joint significance test, and instead presents the results of an ad-hoc "majority rule" count of significant and insignificant individual coefficients.

The standard test of whether or not parcels belong in this equation would be an F test of the null hypothesis that the coefficients of all of the independent variables involving parcels are jointly equal to zero. Doing so using the Postal Service's own regression and statistical procedures, I calculate a P value of 2.7600e-06 in the joint F test, a result that implies rejection of the null hypothesis that variables including parcels do not have any effect on regular delivery time. By the results of this standard test, parcels clearly belong in the regular delivery time equation.

Instead of taking this standard approach Professor Bradley notes that "of the nine terms involving parcels, six of them (66.7 percent) are not significant at the five percent level and five of them are not statistically significant (55.5 percent) at the ten percent level."[46] He then employs

---

[45] NOTICE OF THE UNITED STATES POSTAL SERVICE OF PROVIDING INFORMAL RESPONSES TO UPS QUESTIONS, May 28, 2015, Response to Question 1 (a).

[46] Bradley Report, page 24. His counts appear to be a little off. There are only 8 terms that include parcels, including 5 that are insignificant at the 5 percent level and 4 that are insignificant at the ten percent level. See Initial Neels Report, page 9.

another non-standard test. Rather than dropping the insignificant variables and re-estimating the equation, which would be a commonly accepted response to the presence of insignificant coefficients, he instead picks out the four coefficients that are statistically significant at the 10 percent level from the original regression results and uses them to recompute parcel variability. This non-standard calculation yields a parcel variability of 1.8 percent. However, if instead one excludes the variables with insignificant coefficient, reruns the regression, and recomputes parcel variability, one arrives at a variability of 3 percent, which is very similar to the 2.86 percent implied by my original Modified Proposal Thirteen results.

Professor Bradley also carries out a similar exercise using the national Form 3999 Data Set. He estimates a version of the regular delivery equation that includes my measure of imputed collection mail volume and the DOIS parcel volume. He again pulls a subset of parcel related coefficients out of this model, without re-estimating, and uses them to calculate alternative parcel volume variabilities.[47] He does not comment on whether the parcel related coefficients are jointly significant or have a measurable effect on regular delivery time, a subject that I explore below.

2. Using the National Form 3999 Data to Measure Effects of Parcel volumes on Regular Delivery Time

To test the effect of parcel volumes on regular delivery time, I estimate the Modified Proposal Thirteen model using the National Form 3999 Data Set. Because some variables are missing from the Form 3999 data it is not possible to replicate the proposal thirteen model exactly. To assure the robustness of my findings I test three different specifications. The first of these models uses DOIS parcel volume for the Form 3999 data to the estimate the Modified Proposal Thirteen Model. The second adds both DOIS parcel volume and the imputed collection volume. The third adds the imputed parcel volume (the sum of deviation parcel and in-receptacle parcel) and imputed collection volume.[48] In all cases I follow the spirit of the flexible form, and include the added parcel variable in levels, squares and cross-products. For each version I use the joint F-test of whether parcel volume belongs in each specification. These regression results are shown in Table 11. Extremely small p values for the joint F-test strongly reject the hypothesis that variables including parcels do not have any effect on regular delivery time, regardless of whether

[47] Bradley Report, page 24.

[48] The imputed variables are based on the imputation method described in this report.

the DOIS parcel variable or the imputed parcel variable is used in the model, and regardless of whether collection volumes are included.

Table 11: Modified Proposal Thirteen Estimated Using National Form 3999 Data Set

	DOIS only		DOIS and CV		Imputed PAR and CV	
Variable	Estimated Coefficient	T-statistic	Estimated Coefficient	T-statistic	Estimated Coefficient	T-statistic
INTERCEPT	-7.884	-23.420	-6.705	-18.040	-6.489	-17.030
FSS Dummy	-0.343	-0.537	-0.589	-0.918	0.213	0.316
DPS	0.976	2.973	0.680	1.787	0.972	2.342
DPS2	0.000	-5.599	0.000	-5.019	0.000	-4.886
CM	5.436	9.714	5.004	7.524	5.040	7.448
CM2	0.000	-5.312	0.000	-4.948	0.000	-4.758
SEQ	3.254	6.388	3.161	6.231	1.206	2.100
SEQ2	0.000	-2.742	0.000	-3.389	0.000	-6.783
FSS	10.152	10.080	11.592	9.825	9.180	6.503
CV			-3.708	-3.324	-3.600	-3.195
CV2			0.000	-0.278	0.000	-0.373
DP	19.980	34.170	20.376	30.550	19.224	23.180
DP2	0.000	-7.992	-0.001	-6.716	-0.001	-3.977
DPS*CM	0.000	6.390	0.000	5.904	0.000	7.654
DPS*CV			0.000	1.168	0.000	0.880
DPS*DP	0.000	7.426	0.000	4.802	0.000	2.033
CM*CV			0.000	0.311	0.000	-0.190
CM*DP	0.000	-8.695	-0.001	-6.125	0.000	-2.793
FSS*CV			-0.001	-2.533	-0.001	-1.965
FSS*DP	0.000	-4.806	0.000	-0.564	-0.001	-2.157
CV*DP			0.001	2.006	0.001	1.377
DM	29.180	24.600	28.920	24.620	28.810	24.340
DM2	-18.750	-15.940	-18.510	-15.930	-18.560	-15.850
MPDP	2.404	2.950	1.556	1.971	1.642	2.098
MPDP2	-0.013	-2.908	-0.008	-1.933	-0.009	-2.062
PAR	8.604	0.843	15.552	1.249	26.676	1.876
PAR2	-0.041	-4.354	-0.039	-4.150	-0.148	-2.386
PAR*DPS	0.001	1.568	0.001	1.779	0.001	0.400
PAR*CM	0.002	2.583	0.002	2.621	-0.001	-0.418
PAR*SEQ	-0.002	-1.508	-0.001	-0.975	0.006	5.441
PAR*FSS	0.001	0.747	0.001	0.465	0.010	2.413
PAR*CV			-0.002	-0.416	0.003	0.437
PAR*DP	0.000	-0.289	0.000	-0.277	0.012	2.087
P value on Joint F Test of Parcel Terms	1.04E-10		0		1.45E-08	
R^2	0.935		0.937		0.936	
Observations	10,488		10,488		10,488	

Notes: This table presents the results from of an estimation of Modified Proposal Thirteen on the 3999 Dataset. Consistent with USPS's presentation of its results in

Proposal 13, cost driver coefficients have been converted to seconds while characteristic variable coefficients are expressed in hours. The Form 3999 data does not include data required to calculate business ratios, thus business ratio and its quadratic term are excluded from this estimation.

3. There is No a Priori Reason to Expect Mixed Mail Handling Costs to Be Fixed

Professor Bradley offers a lengthy discussion of the costing implications of situations in which a carrier is simultaneously handling multiple mail streams.[49] He offers a number of examples of situations in which he argues costs are fixed and thus insensitive to changes in mail volume or composition.[50] He concludes by asserting that "the widespread existence of common costs in carrier activities *ensures* (emphasis added) that the simultaneous handling of more than one mail stream by a carrier is not sufficient to justify inclusion of packages in the regular delivery equation."[51]

While I will concede that there may well be mixed mail handling activities whose costs are essentially fixed, I disagree with the implication that in the presence of mixed mail handling activities one needs to prove somehow that parcels belong in the cost modeling equation. To explain my views on this point, let me set aside for the moment the question of the strength or weakness of the non-statistical arguments for including parcel volume in the regular delivery time equation.

In the opening pages of his report Professor Bradley has reaffirmed his belief in the appropriateness of using a flexible form quadratic model for exploring questions of cost causation.[52] The value of such an approach is its ability to accommodate a wide range of potential patterns of cost causation. Suppose that it were in fact the case (as Professor Bradley suggests) that the vast majority of the situations in which a carrier is handling mixed streams of mail that include parcels involve fixed costs that are completely insensitive to changes in mail volume or mix. If so, the flexible quadratic form that Professor Bradley supports would be perfectly capable of accommodating this fact and producing an accurate description of the resulting cost

---

49 Bradley Report, pages 19-20.

50 These include the time required for a carrier to drive from one delivery point to another or the time required for a letter carrier carrying multiples types of mail to deviate from the route's primary line of travel in order to access a customer's receptacle.

51 Bradley Report, page 20.

52 Bradley Report, page 2.

relationships. Suppose in contrast, that the time required to carry out such activities is, in fact, sensitive to the number of parcels included in the mail stream. In that case, excluding parcels from the cost model is likely to yield biased estimates of the remaining model parameters.

From a theoretical standpoint, the choice of whether or not to include parcel variables in a cost causation model is simple. There is no penalty to be paid from including parcels in the cost model if they do not belong; and there is the possibility of biased results and incorrect pricing and regulatory decision making if they belong, but are nonetheless excluded. The obvious conclusion to be drawn from these facts is that one ought to include measures of parcel volume in a regular delivery cost model in order to guard against the possibility of mistakenly attributing parcel costs to other mail products.

Professor Bradley's persistent arguments in support of excluding parcels from the regular delivery time model are all the more remarkable in light of the fact that properly executed statistical tests support their inclusion. I do not believe that citations to 25-year-old testimony describing a dramatically different mail environment are sufficient to offset the statistical results described above.[53]

VII. Conclusions and Recommendations

The comments filed on July 8, 2015 by the Postal Service and other parties heavily criticize the National Form 3999 Model for its reliance on imputed values for parcel and collection mail volumes. At the outset I must emphasize that the use of imputed values cannot by itself serve as a reason for rejecting this model. At the time UPS requested access to the cross walk that would make this analysis possible it made it clear that the development and use of imputed values for these variables were part of the plan. In ordering the Postal Service to make these data available, the Commission implicitly endorsed this plan. Thus, the key issue for the Commission to consider should not be whether the model relies on imputed values, but rather, whether those values and the results based upon them are reliable enough to perform the job they were intended to perform.

The comments offered by Professors Lundblad and Bradley raised a number of issues relating to the construction and performance of these imputed volume variables. In the course of

[53] See Bradley Report, page 22 , footnote 21..

responding to the issues they raised I have made a number of modifications to this analysis that have resulted in measurable improvements. The imputed volume variables are less collinear with other regressors and better able to track volume changes over time. They no longer rely on the DOIS parcel count data whose reliability has been so heavily disputed by the Postal Service. They have passed several rigorous tests based on independently collected nationwide data.

I believe that these changes have improved and strengthened the overall performance of the National Form 3999 Model. I have examined these results closely, and tested them against other available information. I think that the results of this model are reliable and reasonable, and I do not hesitate to urge the Commission to adopt them, recomputing marginal costs and variabilities as needed to reflect target year conditions

I would be less than honest if I did not concede that I would have much preferred to have access to accurate and comprehensive parcel count data for use in this analysis. The development of workable imputed volume measures has been difficult, and there are inherent uncertainties in this process that cannot be entirely eliminated.

Fortunately, this is a fixable problem. There is no reason why the Postal Service cannot start collecting accurate parcel count data as part of the route evaluation process that gives rise to the Form 3999 data. There is no reason why it should not start doing a better job of documenting and tracking the products that it clearly sees as the future of its business. If the Postal Service were to take this step, a valuable resource for improving city carrier cost attribution would emerge rapidly. On average the routes in the Form 3999 Data Set undergo re-evaluation every three years or so. Thus, one year of producing accurate parcel counts as part of the route evaluation process could yield a dataset for updating or re-estimating the National Form 3999 Model that contained over 3,000 zip code level observations – far more than the number involved in the special parcel study.

If the Commission is not ready to adopt the results of the National Form 3999 model, I would continue to urge that it embrace what I have called the Modified Proposal Thirteen. It seems entirely plausible to me that the presence of large volumes of parcels might have some effect on what the Postal Service calls regular delivery time. More importantly, it also seems reasonable to UPS personnel, who have far more familiarity with these types of operations than I do. If the Postal Service were correct in its assertions that parcels have no effect on regular delivery time we should find that measures of parcel volume have no effect when added to econometric models of regular delivery. But, that is not what we find. Instead we find that measures of parcel

volumes have positive and statistically significant effects when added either to the econometric model developed in connection with Proposal Thirteen, or to an analogous model based on the Form 3999 dataset. These are hard facts to explain away. The Commission should feel completely comfortable adopting this proposal.